

MEM第二次作业

注：

- 回答使用中英文皆可
 - 推荐使用Rmd或者其他支持markdown的书写工具（如免费工具MarkText，收费Typora)答题。
 - 请在github里提交你的作业
-

Question #1: BigBangTheory. (Attached Data: BigBangTheory)

The Big Bang Theory, a situation comedy featuring Johnny Galecki, Jim Parsons, and Kaley Cuoco-Sweeting, is one of the most-watched programs on network television. The first two episodes for the 2011–2012 season premiered on September 22, 2011; the first episode attracted 14.1 million viewers and the second episode attracted 14.7 million viewers. The attached data file BigBangTheory shows the number of viewers in millions for the first 21 episodes of the 2011–2012 season (*the Big Bang theory* website, April 17, 2012).

- Compute the minimum and the maximum number of viewers.
- Compute the mean, median, and mode.
- Compute the first and third quartiles.
- has viewership grown or declined over the 2011–2012 season? Discuss.

Question #2: NBAPlayerPts. (Attached Data: NBAPlayerPts)

CbSSports.com developed the Total Player Rating system to rate players in the National Basketball Association (NBA) based on various offensive and defensive statistics. The attached data file NBAPlayerPts shows the average number of points scored per game (PPG) for 50 players with the highest ratings for a portion of the 2012–2013 NBA season (CbSSports.com website, February 25, 2013). Use classes starting at 10 and ending at 30 in increments of 2 for PPG in the following.

- Show the frequency distribution.
- Show the relative frequency distribution.
- Show the cumulative percent frequency distribution.
- Develop a histogram for the average number of points scored per game.

e. Do the data appear to be skewed? Explain.

f. What percentage of the players averaged at least 20 points per game?

Question #3: A researcher reports survey results by stating that the standard error of the mean is 20. The population standard deviation is 500.

a. How large was the sample used in this survey?

b. What is the probability that the point estimate was within ± 25 of the population mean?

Question #4: Young Professional Magazine (Attached Data: Professional)

Young Professional magazine was developed for a target audience of recent college graduates who are in their first 10 years in a business/professional career. In its two years of publication, the magazine has been fairly successful. Now the publisher is interested in expanding the magazine's advertising base. Potential advertisers continually ask about the demographics and interests of subscribers to *young Professionals*. To collect this information, the magazine commissioned a survey to develop a profile of its subscribers. The survey results will be used to help the magazine choose articles of interest and provide advertisers with a profile of subscribers. As a new employee of the magazine, you have been asked to help analyze the survey results.

Some of the survey questions follow:

1. What is your age?

2. Are you: Male_____ Female_____

3. Do you plan to make any real estate purchases in the next two years?

Yes_____ No_____

4. What is the approximate total value of financial investments, exclusive of your home, owned by you or members of your household?

5. How many stock/bond/mutual fund transactions have you made in the past year?

6. Do you have broadband access to the Internet at home? Yes_____ No_____

7. Please indicate your total household income last year. ____

8. Do you have children? Yes_____ No_____

The file entitled Professional contains the responses to these questions.

Managerial Report:

Prepare a managerial report summarizing the results of the survey. In addition to statistical summaries, discuss how the magazine might use these results to attract advertisers. You might also comment on how the survey results could be used by the magazine's editors to identify topics that would be of interest to readers. Your report should address the following issues, but do not limit your analysis to just these areas.

- a. Develop appropriate descriptive statistics to summarize the data.
- b. Develop 95% confidence intervals for the mean age and household income of subscribers.
- c. Develop 95% confidence intervals for the proportion of subscribers who have broadband access at home and the proportion of subscribers who have children.
- d. Would *Young Professional* be a good advertising outlet for online brokers? Justify your conclusion with statistical data.
- e. Would this magazine be a good place to advertise for companies selling educational software and computer games for young children?
- f. Comment on the types of articles you believe would be of interest to readers of *Young Professional*.

Question #5: Quality Associate, Inc. (Attached Data: Quality)

Quality associates, inc., a consulting firm, advises its clients about sampling and statistical procedures that can be used to control their manufacturing processes. In one particular application, a client gave Quality associates a sample of 800 observations taken during a time in which that client's process was operating satisfactorily. The sample standard deviation for these data was .21; hence, with so much data, the population standard deviation was assumed to be .21. Quality associates then suggested that random samples of size 30 be taken periodically to monitor the process on an ongoing basis. By analyzing the new samples, the client could quickly learn whether the process was operating satisfactorily. When the process was not operating satisfactorily, corrective action could be taken to eliminate the problem. The design specification indicated the mean for the process should be 12. The hypothesis test suggested by Quality associates follows.

$$H_0 : \mu = 12$$

$$H_1 : \mu \neq 12$$

Corrective action will be taken any time H_0 is rejected.

Data are available in the data set Quality.

Managerial Report

- a. Conduct a hypothesis test for each sample at the .01 level of significance and determine what action, if any, should be taken. Provide the p-value for each test.
- b. compute the standard deviation for each of the four samples. does the assumption of .21 for the population standard deviation appear reasonable?
- c. compute limits for the sample mean \bar{x} around $\mu = 12$ such that, as long as a new sample mean is within those limits, the process will be considered to be operating satisfactorily. if \bar{x} exceeds the upper limit or if \bar{x} is below the lower limit, corrective action will be taken. these limits are referred to as upper and lower control limits for quality control purposes.
- d. discuss the implications of changing the level of significance to a larger value. what mistake or error could increase if the level of significance is increased?

Question #6: Vacation occupancy rates were expected to be up during March 2008 in Myrtle Beach, South Carolina (*the sun news*, February 29, 2008). Data in the file Occupancy (Attached file **Occupancy**) will allow you to replicate the findings presented in the newspaper. The data show units rented and not rented for a random sample of vacation properties during the first week of March 2007 and March 2008.

- a. Estimate the proportion of units rented during the first week of March 2007 and the first week of March 2008.
- b. Provide a 95% confidence interval for the difference in proportions.
- c. On the basis of your findings, does it appear March rental rates for 2008 will be up from those a year earlier?

Question #7: Air Force Training Program (data file: Training)

An air force introductory course in electronics uses a personalized system of instruction whereby each student views a videotaped lecture and then is given a programmed instruction text. the students work independently with the text until they have completed the training and passed a test. Of concern is the varying pace at which the students complete this portion of their training program. Some students are able to cover the programmed instruction text relatively quickly, whereas other students work much longer with the text and require additional time to complete the course. The fast students wait until the slow students complete the introductory course before the entire group proceeds together with other aspects of their training.

A proposed alternative system involves use of computer-assisted instruction. In this method, all students view the same videotaped lecture and then each is assigned to a computer terminal for

further instruction. The computer guides the student, working independently, through the self-training portion of the course.

To compare the proposed and current methods of instruction, an entering class of 122 students was assigned randomly to one of the two methods. one group of 61 students used the current programmed-text method and the other group of 61 students used the proposed computer-assisted method. The time in hours was recorded for each student in the study. Data are provided in the data set training (see Attached file).

Managerial Report

- a. use appropriate descriptive statistics to summarize the training time data for each method. what similarities or differences do you observe from the sample data?
- b. Comment on any difference between the population means for the two methods. Discuss your findings.
- c. compute the standard deviation and variance for each training method. conduct a hypothesis test about the equality of population variances for the two training methods. Discuss your findings.
- d. what conclusion can you reach about any differences between the two methods? what is your recommendation? explain.
- e. can you suggest other data or testing that might be desirable before making a final decision on the training program to be used in the future?

Question #8: The Toyota Camry is one of the best-selling cars in North America. The cost of a previously owned Camry depends upon many factors, including the model year, mileage, and condition. To investigate the relationship between the car's mileage and the sales price for a 2007 model year Camry, Attached data file Camry show the mileage and sale price for 19 sales (Pricehub website, February 24, 2012).

- a. Develop a scatter diagram with the car mileage on the horizontal axis and the price on the vertical axis.
- b. what does the scatter diagram developed in part (a) indicate about the relationship between the two variables?
- c. Develop the estimated regression equation that could be used to predict the price (\$1000s) given the miles (1000s).
- d. Test for a significant relationship at the .05 level of significance.

- e. Did the estimated regression equation provide a good fit? Explain.
- f. Provide an interpretation for the slope of the estimated regression equation.
- g. Suppose that you are considering purchasing a previously owned 2007 Camry that has been driven 60,000 miles. Using the estimated regression equation developed in part (c), predict the price for this car. Is this the price you would offer the seller.

Question #9: 附件WE.xlsx是某提供网站服务的Internet服务商的客户数据。数据包含了6347名客户在11个指标上的表现。其中“流失”指标中0表示流失，“1”表示不流失，其他指标含义看变量命名。

- a. 通过可视化探索流失客户与非流失客户的行为特点（或特点对比），你能发现流失与非流失客户行为在哪些指标有可能存在显著不同？
- b. 通过均值比较的方式验证上述不同是否显著。
- c. 以“流失”为因变量，其他你认为重要的变量为自变量（提示：a、b两步的发现），建立回归方程对是否流失进行预测。
- d. 根据上一步预测的结果，对尚未流失（流失=0）的客户进行流失可能性排序，并给出流失可能性最大的前100名用户ID列表。